Economics 102
Homework 3

Department of Economics — **Stanford University**
Due: October 21, 1999 — **Professor Frank A. Wolak**

1. The Las Vegas casinos run a gambling market for football game outcomes. In this market, participants place bets on either side of a point spread. The point spread for a football game is defined as the estimated difference between the points scored by the home team and the away team in a football game. Suppose that San Francisco 49ers are playing the Oakland Raiders at 3COM park and the casino sets the point spread at -3. This means that a person betting on the 49ers will win their bet if the 49ers win the game or lose the game by less than 3 points. A person betting of the Oakland Raiders will win their bet if the Raiders win by more than 3. If the Raiders win by 3 points, then no one wins and all bets are returned. The casinos make money by operating this market. They take approximately $1 of every $11 won in vigorish or "vig" as the market participants call it.

Suppose you would like to enter this market. You have a dataset of final Friday (before the Sunday or Monday game) Points Spread (PS) and actual game outcome Point Differences (PD) for all games played from 1983 to 1985. You have two goals for this problem. The first is to investigate the validity of the null hypothesis that the "Vegas Line" point spread, PS, is an unbiased estimate of the actual point difference in the game. The second is to find a linear regression model based on data based on actual team performance–yards rushing, yards passing, fumbles, takeaways–to predict actual point difference, PD, between two teams playing in order to devise a profitable betting strategy. To this end you will be given data for three years, from 1983 to 1985 to formulate your betting strategy model. Complete the following tasks.

(a) For each year of data run the regression: PD(i) = a(j) + b(j)PS(i) + e(i), where PD(i) is the point difference for game i, PS(i) is the Vegas Line point spread for game i, and e(i) is linear regression disturbance. The coefficients a(j) and b(j) are indexed by j, the year, to indicate that they can be differ across the three years.

(b) For each year test the joint null hypothesis, H: a=0 and b=1 versus K: either $a \neq 0$ or $b \neq 1$using a size $\alpha = 0.01$ test. (Using this size for all remaining tests performed in this problem.)

(c) Combine all three years of data and estimate PD(i) = a + bPS(i) + e(i), for all three years of data jointly. Test the null hypothesis H: a(1) = a(2) = a(3) and b(1) = b(2) = b(3) versus the alterative that at least one of these equalities does hold.

(d) Using the effort variables these teams from previous weeks estimate a linear regression model to predict the point difference (PD) for the game between the two teams using the following regressors for the three years of data :

OFFHOME = the number of offensive yards gained last week by the home team,
DEFHOME = the number of defensive yards allowed last week by the home team,
TAKEHOME = the number of takeaways last week by the home team,
OFFAWAY = the number of offensive yards gained last week by the away team,
DEFAWAY = the number of defensive yards allowed last week by the away team,
TAKEAWAY = the number of takeaways last week by the away team

Be sure to include a constant term in your regression. You will have to drop the first week of data for each year from your sample. Test the null hypothesis that coefficients associated with all of these variables besides the constant term are jointly zero, versus the alternative that at least one coefficient is not zero.

(e) Re-estimate the model in (d) including PS as an additional regressor and test the joint null hypothesis that all other coefficients in the model besides PS are equal to zero versus the alternative that at least one of these coefficients is non-zero.

2. Consider the model

$$y_t = \beta_1 + x_{2t}\,\beta_2 + x_{3t}\,\beta_3 + e_t$$

and suppose that application of least squares to 20 observations on these variables yields the following results

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.965 \\ 0.699 \\ 1.777 \end{bmatrix} \qquad \hat{cov}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.218 & 0.0191 & -0.050 \\ 0.0191 & 0.048 & -0.031 \\ -0.050 & -0.031 & 0.037 \end{bmatrix}$$

$$s^2 = 2.519 \qquad R^2 = 0.9466$$

(a) Find the total sum of squares, the residual sum of squares and explained sum of squares for this model.

(b) Find 95 percent interval estimate for β_2 and β_3.

(c) Use a t-test to test the hypothesis H: $\beta_2 \geq 1$ against the alternative K: $\beta_2 < 1$.

(d) Use your answer in part (a) to test the joint hypothesis H: $\beta_2 = 0$, $\beta_3 = 0$ versus the alternative that at least one of the equalities does not hold.

3. In the model

$$y_t = \alpha + x_t\,\beta + u_t$$

x_t is a non-random exogenous variable and the u_t are serially uncorrelated random disturbances with zero mean and variance σ^2 for each value of t. The following sample moments have been calculated from 10 observations on y_t and x_t

$$\sum_{t=1}^{T} y_t = 8, \quad \sum_{t=1}^{T} x_t = 40, \quad \sum_{t=1}^{T} y_t^2 = 26, \quad \sum_{t=1}^{T} x_t^2 = 200, \quad \sum_{t=1}^{T} x_t y_t = 20,$$

where T= 10. In some subsequent time period, s, for which the above regression model is still valid the value of $x_s = 10$.

(a) Calculate the best linear unbiased forecast of y_s using the above data.

(b) Estimate the standard error of your forecast in (a).